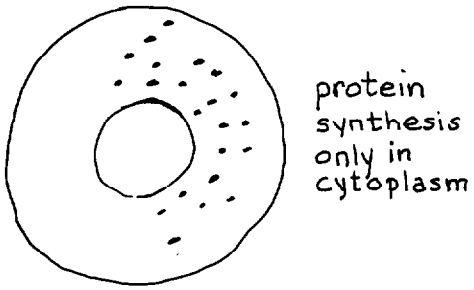
Lecture 12

Protein synthesis

In the last two lectures we have seen how cells make DNA and RNA. Now we come to the synthesis of protein. Again, the cell uses two devices with which we are already familiar: (1) a set of enzymes to activate the precursor substances--the amino acids--and put them together to form protein chains; and (2) a template. In protein synthesis, in addition, there is one more problem: the information of the genes, which is present in DNA or RNA as a sequence of nucleotides, has to be translated from the language of the nucleic acids into the language of the proteins, a sequence of amino acids. There must, therefore, exist specific rules of translation and these are called the genetic code. To decipher the secrets of the genetic code required a combination of genetic and biochemical techniques. We shall begin with the biochemical approaches to protein synthesis. Then after a discussion of the genetic code, there will be a gap in the story until we come back to it with some knowledge of genetics.

Protein synthesis can be measured by monitoring the incorporation of labeled amino acids into polypeptides, which are acid insoluble. Is the protein made in any special place in the cell? Bacteria are too small for us to localize the sites of protein synthesis within them. In eucaryotic cells, protein synthesis is easily seen to occur in the cytoplasm: a short labeling with radioactive amino acids followed by autoradiography shows the label localized only on the cytoplasm. Only later some proteins return to the



nucleus. Since practically all the DNA in the cell is located in the nucleus it is clear that protein is not made directly on or near the DNA. It is made in the cytoplasm using information that comes from the DNA via messenger RNA. The

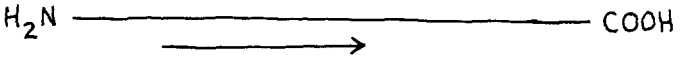
mechanism that transports mRNA from the nucleus where it is made to the cytoplasm where it is used is still unknown. There are indications that the messenger gets processed by enzymes that chew off some parts of it.

Direction of synthesis

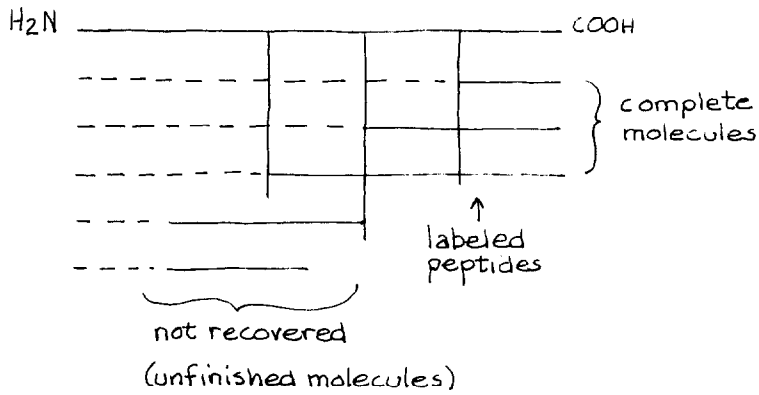
Can one tell, from experiments on intact cells how a polypeptide chain is made? As you know, each polypeptide chain contains an amino group at one end and a carboxyl group at the other end. The chains might be made in several ways; for example:

1. in many pieces that are then stuck together;
2. in one piece starting from the carboxyl end;
3. ditto but starting from the amino end; or
4. starting from the middle and growing toward the two ends.

The answer is simple: all polypeptide chains are made starting at the end with the terminal amino group and ending at the terminal carboxyl group.



This was shown in 1960 by Howard Dintzis, who used reticulocytes, that is, immature red blood cells from rabbits. (Mature red blood cells have stopped making protein and cannot be used.) Reticulocytes make one protein only, hemoglobin. This protein, whose properties we shall consider later in some detail, consists of four polypeptide chains, a pair of α chains and a pair of β chains: $\alpha_2\beta_2$. (To each chain there is attached a nonprotein heme group.) The bone marrow of a rabbit can be stimulated to produce reticulocytes in large excess and release them into the bloodstream. In Dintzis' experiments, a reticulocyte suspension was used to label growing hemoglobin chains, which were then isolated from complete hemoglobin molecules. The label was the amino acid leucine, of which there are many copies all along the α and the β chains. Dintzis first found that it took about one minute to make a complete β chain. Then he added to a suspension of reticulocytes the labeled leucine for a short period, about 30 seconds; stopped protein synthesis by chilling the suspension; isolated the complete $\alpha_2\beta_2$ hemoglobin molecules; purified them; separated the α and β chains. Then he broke the β chains into specific



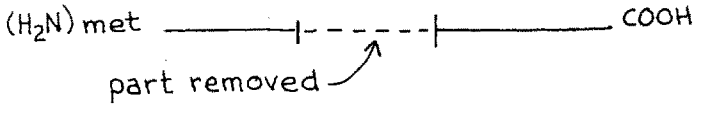
fragments (or peptides) by means of trypsin, separated the peptides by electrophoresis and chromatography, and asked how the label was distributed into various regions of the chains. Where would one expect to find label in finished hemoglobin chains? Primarily in those portions of the chains that are made last since the labeling had lasted only one-half the time needed to make a chain and since what was examined was the chains that had been finished. What Dintzis found was that the completed hemoglobin chains were labeled predominantly near the carboxyl end, and the closer a fragment was to that end, the more labeling it had. The peptides near the H_3N^+ end were unlabeled.

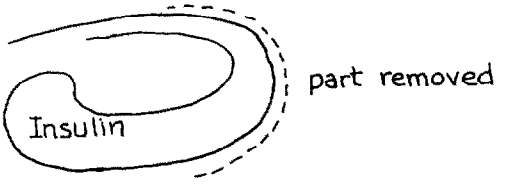
In the reverse experiment labeled leucine was added for a short period of time and then a hundredfold excess of nonlabeled leucine was added to swamp out the radioactive label. Then some time was allowed to pass before the reaction was stopped. In this pulse-chase experiment, more label was found present at the amino terminus because many partly labeled molecules could be completed in unlabeled medium. [Notice that the key to interpreting these experiments was that only finished hemoglobin molecules were analyzed. Unfinished molecules would have contributed labels in all parts of the chain if they had been examined along with the finished ones.]

Dintzis' experiments were done in 1960, before the discovery of techniques for synthesizing proteins in vitro. Since then it has been confirmed that all protein chains are made in the same way, starting at the amino end and going toward the carboxyl end.

Do all proteins start with the same amino acid? There are reasons to expect that this is so. If the signal that directs the starting point for making protein at a particular site on a messenger RNA were a unique sequence representing a given amino acid, all proteins in a given organism might be expected to begin with the same amino acid. If not, then there must be some other signal that says "start here" but is not itself translated into protein. (Think of a bus stop signal that directs a bus to stop 100 feet down the road.) It seems now fairly certain that all proteins in all organisms that have been tested start with the same amino acid, either methionine or a modified form of methionine called formyl methionine. Yet, when proteins are extracted from E. coli cells, methionine is the most frequent but by no means the only amino acid found at the free amino end.

The explanation is that, as polypeptide chains are made, one or more amino acids may be snipped off one or the other end or even from the middle. For example, insulin is first made as proinsulin, a chain of 84 amino acids from which certain enzymes cut out an internal piece 33 amino acids long. Only then does insulin become active.



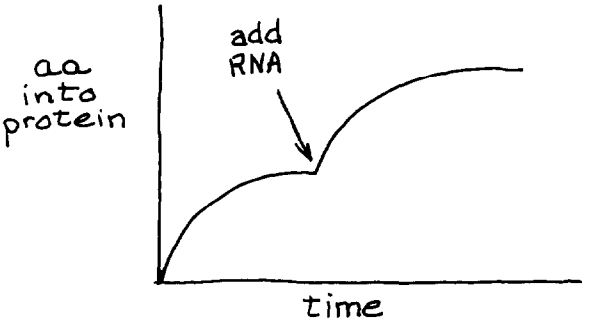


We know that protein chains in general begin with an amino terminal methionine (or formyl methionine) because we can now study the synthesis of proteins in the test tube using extracts of bacteria, plant, or animal cells.

In vitro synthesis

The key to success was discovered by Nirenberg in 1961. Most cell extracts incorporate amino acids into protein for only a few minutes.

Nirenberg found that, after an extract from E. coli had apparently died, protein synthesis would restart if he added some RNA to it.

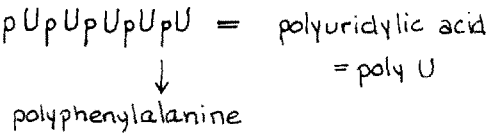


This was an enormously important discovery. It made the study of protein synthesis and its regulation amenable to the methods of enzymology. Biochemists like to work with extracts in test tubes rather than with intact live cells. With Nirenberg's system it was possible to find out what in the extract was needed to make protein. There had to be ribosomes, transfer RNA's (the adapter molecules to which the amino acids must be attached), the amino acids, of course, and the activating enzymes that attach each amino acid to its adapter tRNA. In addition the extract contains some proteins that play various roles in the complicated process of making protein. ATP and GTP are needed to provide energy for specific steps. Finally a messenger RNA has to be present. [The reason extracts from E. coli stopped making protein is that bacterial mRNA is unstable: it is chewed up after it has been used to make a few dozen protein molecules. The mRNA of animal and plant cells lasts much longer. Can you think of some good reason for this stability? Think of the need for rapid adaptation versus stable differentiation.]

In Nirenberg's extracts, only mRNA served as template. Ribosomal RNA or tRNA did not serve as messengers: they do not code for protein. Instead, the RNA from viruses was a good messenger and directed the synthesis of the same proteins that are found in the corresponding virions. This was a most gratifying result indeed.

Even more exciting, Nirenberg found that artificially made polynucleotides could serve as messengers, although rather poorly: Each sequence

caused incorporation of specific amino acids: for example, polyphenylalanine was made under the direction of poly U. The genetic code was born.

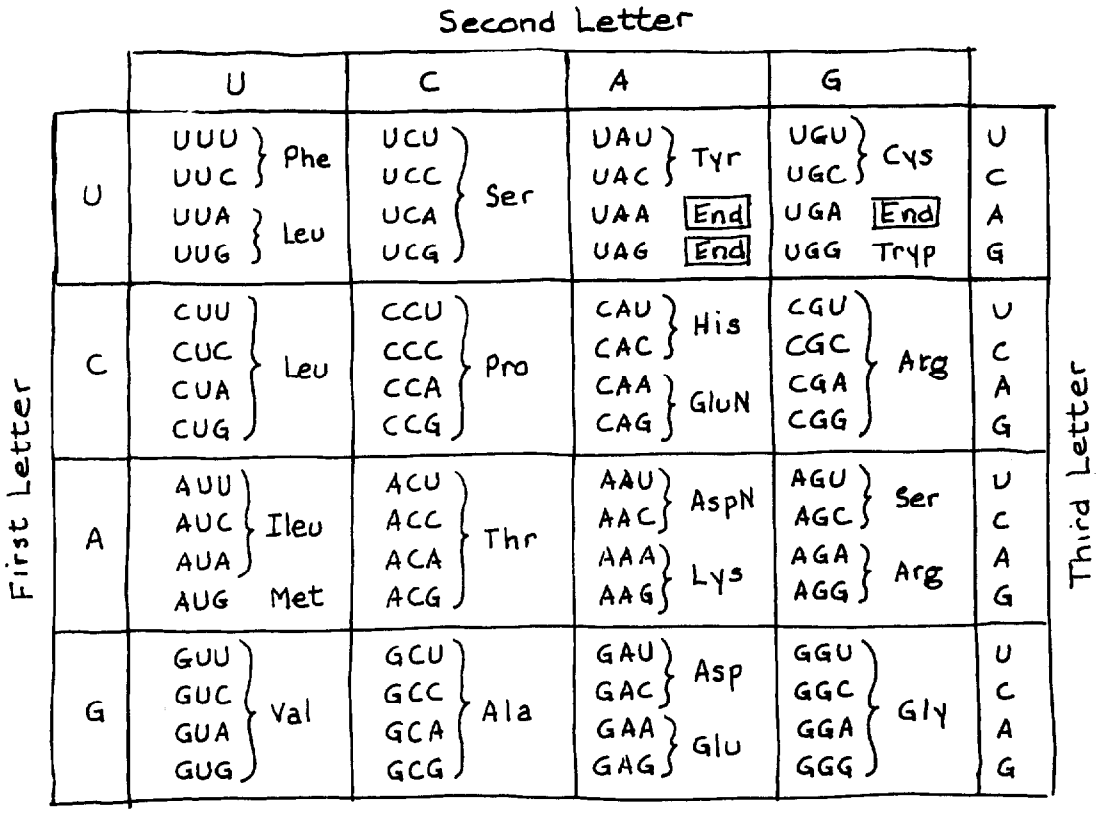


The necessary existence of rules of translation between the alphabet of nucleic acids and that of protein had, of course, been discussed before Nirenberg's experiments, but only with his system did it become possible to find out "what meant what."

Genetic code

We'll outline now that knowledge although several ends will be left dangling until we discuss molecular genetics. The genetic code is a triplet code: three adjacent nucleotides specify an amino acid. For example, pUpUpU (or UUU for short) means polyphenylalanine. Mixed polynucleotides such as ...CUCUCU... in Nirenberg's extracts code for incorporation of mixtures of amino acids into protein.

The idea of a triplet code is a natural consequence of the fact that 4 nucleotides either one-by-one or two-by-two ($4^2 = 16$) do not provide enough combinations to code for 20 amino acids. Three-by-three ($4^3 = 64$) can work. In fact, it turns out to be just so, and every one of the 64 possible triplets is used, either as an amino acid codon or as a signal for termination.



But what decides where a triplet starts? Are there "commas"? No. The code is comma-free; there are only start signals, and the reading proceeds down the messenger three nucleotides by threes in a fixed direction from the start point. That is why poly U or poly A, etc., all synthetic nucleotides, are lousy messengers: they lack the start signals.

Universality

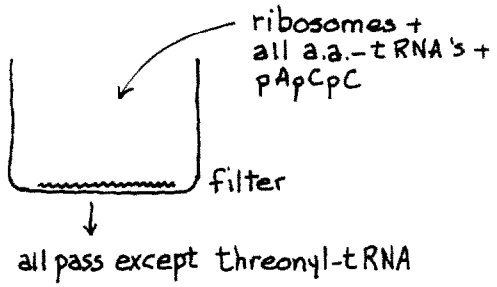
The genetic code is universal or nearly so. If you feed to a Nirenberg extract from E. coli an RNA messenger from tobacco mosaic virus or from human blood cells (or vice versa, a messenger from E. coli to a mammalian cell extract) the protein that is made is the same as would be made in the cell of origin of the messenger. This means that the translation dictionary is the same: the codons are the same and the tRNA's recognize the same codons to insert the same amino acids.

Ever since cellular life appeared perhaps 2 billion years ago, nothing much has changed in the genetic code. The enzymes that attach the amino acid to the tRNA must of course have evolved as DNA evolved; but the recognition of an mRNA signal--

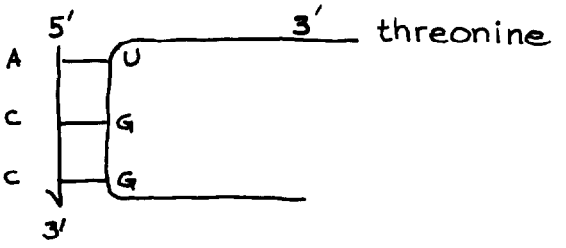
a codon--by the tRNA that carries a specific amino acid has apparently remained constant throughout the whole period. There are many reasons why this should be so. Imagine a mutation that changed the recognition of a codon: it would affect not one protein, but practically all proteins of the organism. The chances that such a mutation is not lethal are almost nil. This conclusion as to the universality and stability of the genetic code is on a par with the finding that genetic material in every organism is composed of nucleic acid and functions through its sequence of nucleotides. It places together all organisms, past and present, not only in genetic structure but in mechanism of genetic expression.

The individual codon assignments were decided on the basis of experiments of many different types. Certain chemicals can act directly on RNA and change one nucleotide into another in a chemically defined way. For example, nitrous acid acting on RNA converts cytosine to uracil (C to U). If this treatment is used on an RNA that makes a well known protein--the classical case is the RNA of tobacco mosaic virus--the altered RNA makes an altered protein with specific amino acid substitutions. These provide information of codon changes due to C → U changes.

The most elegant confirmation of the codon chart came when Nirenberg and his colleagues discovered that a free triplet, for example, the trinucleotide pApCpC (or ACC), while of course not causing incorporation of any amino acids into protein (it has only one codon!) causes the appropriate amino acyl tRNA (aa-tRNA) to stick specifically to ribosomes. If it is added to a mixture of all aa-tRNA's plus ribosomes which is then filtered, ACC causes only specific threonyl-tRNA to remain on a filter, while the other amino acids go right through.



All codon assignemnts have been verified in the same way. Evidently, a little RNA piece such as ACC can hold together ribosome and the amino acyl tRNA strongly enough to keep the latter from going through the filter. The codon pairs with the anticodon (for example, UGG if the codon is ACC) in the tRNA by the usual 2 or 3 hydrogen bonds per nucleotide.



Degeneracy

Of the 64 codons, 61 code for amino acids: the code is degenerate (that is, it provides more than one solution for a problem). The extent of degeneracy varies: for example, there are six different codons for leucine and only one for methionine. For each codon on the messenger, of course, there must be at least one transfer RNA to bring that amino acid to that position.

Further reading

Watson's Chapters 12, "Involvement of RNA in Protein Synthesis," and 13, "The Genetic Code," as well as Stent's Chapter 17, "RNA Translation," are useful supplements to this lecture and the next (lecture 13).